

Serial Number: 10/042,065A

ENTERED

- ☐ Changed a file from non-ASCII to ASCII
- ☐ Changed the margins in cases where the sequence text was wrapped down to the next line.
- ☐ Edited a format error in the Current Application Data section, specifically: _____
- ☐ Edited the Current Application Data section with the actual current number. The number inputted by the applicant was ☐ the prior application data; or ☐ other _____
- ☐ Added the mandatory heading and subheadings for "Current Application Data".
- ☐ Edited the "Number of Sequences" field. The applicant spelled out a number instead of using an integer.
- ☐ Changed the spelling of a mandatory field (the headings or subheadings), specifically: _____
- ☐ Corrected the SEQ ID NO when obviously incorrect. The sequence numbers that were edited were: _____
- ☐ Inserted or corrected a nucleic number at the end of a nucleic line. SEQ ID NO's edited: _____
- ☐ Corrected subheading placement. All responses must be on the same line as each subheading. If the applicant placed a response below the subheading, this was moved to its appropriate place.
- ☐ Inserted colons after headings/subheadings. Headings edited included: _____
- ☐ Deleted extra, invalid, headings used by an applicant, specifically: _____
- ☒ Deleted: ☒ non-ASCII "garbage" at the beginning/end of files; ☐ secretary initials/filename at end of file; ☐ page numbers throughout text; ☐ other invalid text, such as _____
- ☐ Inserted mandatory headings, specifically: _____
- ☐ Corrected an obvious error in the response, specifically: _____
- ☐ Edited identifiers where upper case is used but lower case is required, or vice versa.
- ☐ Corrected an error in the Number of Sequences field, specifically: _____
- ☐ A "Hard Page Break" code was inserted by the applicant. All occurrences had to be deleted.
- ☐ Deleted *ending* stop codon in amino acid sequences and adjusted the "(A)Length:" field accordingly (error due to a PatentIn bug). Sequences corrected: _____
- ☐ Other: _____

*Examiner: The above corrections must be communicated to the applicant in the first Office Action. DO NOT send a copy of this form.



OIPE

RAW SEQUENCE LISTING

DATE: 07/11/2002

PATENT APPLICATION: US/10/042,665A

TIME: 09:11:59

Input Set : A:\ptoms.txt

Output Set: N:\CRF3\07112002\J042665A.raw

```

3 <110> APPLICANT: Schupp, Thomas
4     Toupet, Christine
5     Engel, Nathalie
7 <120> TITLE OF INVENTION: Rifamycin biosynthesis gene cluster
9 <130> FILE REFERENCE: 4-21001/B/C1
11 <140> CURRENT APPLICATION NUMBER: 10/042,665A
12 <141> CURRENT FILING DATE: 2002-01-09
14 <150> PRIOR APPLICATION NUMBER: 09/242,744
15 <151> PRIOR FILING DATE: 1999-03-24
17 <150> PRIOR APPLICATION NUMBER: PCT/EP97/04495
18 <151> PRIOR FILING DATE: 1997-08-18
20 <160> NUMBER OF SEQ ID NOS: 9
22 <170> SOFTWARE: PatentIn Ver. 2.1
24 <210> SEQ ID NO: 1
25 <211> LENGTH: 5676
26 <212> TYPE: DNA
27 <213> ORGANISM: Amcylatopsis mediterranei
29 <400> SEQUENCE: 1
30 ggtacccggt gttcgcgacg gcgttcgacg aggcttgcca gcagctggac gtctgtctgg 60
31 ccggccgtgc cgggcaaccg gtgcgggacg tcgtgctcgg cgaagtgcc gccgaaaccg 120
32 ggctgttgaa ccagacggtc ttcacccaag ccgggctgtt cgcggtggag agcgcgctgt 180
33 tccggtctgc cgaatcctgg ggtgtccggc cggacgtggt gtcgcccac tccatcgggg 240
34 agatcaccgc cgcgtatgcc gcgggcgtct tctcgtcgcc ggacgcccgc cggatcgctc 300
35 cggcgcgagg ccggctgatg caggcgctgg cggcgggcgg ggcgatggtc gccgtcgccg 360
36 cctccgaagc cgaggtggcc gaactgctcg gcgacggcgt ggaactcgcc gccgtcaacg 420
37 gcccttcggc ggtagtcctt tccggggacg cggacgcggg cgtcgcgggc gccgcccgcg 480
38 tgcgcgagcg cgggcacaag accaagcagc tcaaggtttc gcacgcgttc cactccgcgc 540
39 ggatggcgcc gatgctggcg gagttcgccg ccgagctggc cggcgtagcg tggcgcgagc 600
40 cggagatccc ggtggtctcc aacgtgaccg gccggttcgc cgagcccggc gaactgaccg 660
41 agccgggcta ctgggcccag cacgtgcggc ggccggtgag gttcgccgag ggcgtcgcg 720
42 ccgcgacgga gtccggcggc tcgtgttctg tggagctcgg gccgggggag gcgctgaccg 780
43 ccctcgctga ggagacggcc gaggtcacct gcgtcgcggc cctgcgggac gaccgcccgg 840
44 aggtcaccgc gctgatcacc gcggtcgccg agctgttcgt ccgcgggggt gcggtcgatt 900
45 ggccggccct gctgcccggc gtcaccgggt tcgtcgacct gccgaagtac gccttcgacc 960
46 agcagcacta ttggctgcag cccgcccgcg aggccacgga cgcggcctcg ctccggcgag 1020
47 tcgcggccga ccaccgcgtg ctggggcgcg tgggtccggc gccgcagtcg gacggcctgg 1080
48 tcttcacctc gcggtgtgca ttgaaatcgc acccggtggc ggccgaccac gtcacgcgcg 1140
49 ggggtggtgt cgtcgcgggc accgggctcg tcgagctggc cgtccggggc ggggacgagg 1200
50 ccggtgccc ggtcctcgaa gaactcgtca tcgaggtccc gctggtcgtc cccgaccacg 1260
51 gcgggggtcc gatccaggtc gtcgtggggg caccggggga gaccggttcg cgcgcggtcg 1320
52 aggtgtactc cctgcgcgag gacgcccgtg ccgaagtgtg ggcccggcac gccaccgggt 1380
53 tcttggtcgc gacgcccgtc cagcacaagc cgttcgactt caccgcctgg ccgcccggcg 1440
54 gcgtcgagcg cgtcgacgtc gaggacttct acgacggcct cgtcgaccgc gggtagcct 1500

```

RAW SEQUENCE LISTING

DATE: 07/11/2002

PATENT APPLICATION: US/10/042,665A

TIME: 09:11:59

Input Set : A:\ptoms.txt

Output Set: N:\CRF3\07112002\J042665A.raw

```

55 acgggcccgtc gttccggggc ctgcccggcg tgtggcggcg cggcgacgaa gtgttcgccg 1560
56 aggtcgccct ggccgaggac gaccgcgcgg acgcggcccg gttcggcatc caccgccgcc 1620
57 tgctggacgc cgcctgcac gcgggcatgg ccggtgccac caccacgga gagcccgcc 1680
58 ggccggtgct gccgttcgcc tggaaacggc tgggtgtgca cgcggccggg gcgtccgcgc 1740
59 tgcgggtccg gctcgccccg agcgggtccg acgcccgtgc ggtcgaggcc gcggacgagg 1800
60 ccggcgggtct cgttgtgacg gcggactcgc tgggtctccc gccgggtgtc gccgaacagc 1860
61 tgggcgcggc ggccgaaccac gacgcgttgt tccgcgtgga gtggaccgag atttcctcgg 1920
62 ctgggagacgt tccggcggac cagctcgaag tgctcgaagc cgtcggcgag gatcccctgg 1980
63 aactgaccgg ccgggtcctg gaggccgtgc agacctggct cgcgcacgca gccgacgacg 2040
64 ctgcctggt cgtggtgacc cgcggcgccg tccacgaggt gactgaccgg gccggtgccg 2100
65 cgtgtgggg cctgatccgg gccgcgcagg cggaaaaccc ggaccggatc gtgtgtctgg 2160
66 acaccgacgg tgaagtgcg ctaggcgggg tgctggccac cggcgagccc caaacagccg 2220
67 tccgaggcgc cagctgttc gccccgcggc tggcccgcgc cggaggccgc gaggcaccgg 2280
68 cagtgaccgg cgggacggtc ctgatctcgg gcgcgcgtc gctgggcgcg ctaccgccc 2340
69 ggcacctggt cgcgcggcac ggagtcgcgg ggtgtgtgct cgtcagccgc cgtggccccg 2400
70 acgcccacgg catggccgaa ctgaccgctg aactcatcgc tcagggcgcc gaggtcgccg 2460
71 tagtcgcttg cgacctggcc gaccgggacc aggtccgggt actgctggcc gagcaccgcc 2520
72 cgaacgccgt cgtgcacacg gccggtgttc tcgacgacgg cgtcttcgag tcgctgacgc 2580
73 gggagcggct ggccaaggtc ttgcgcacca aagttactgc tgccaatcac ctcgacgagc 2640
74 tgaccgcga actggatctt cgcgcgttcg tegtgttctc ctccgcctcc ggggtcttcg 2700
75 gctccgcgg gcagggaac tacgcgcgtg ccaacgccta cctggacgcc gtggtcgcca 2760
76 accgcccggc cgcgggcctg cccggcacat cgtgggcctg gggcctgtgg gaacagaccg 2820
77 acgggatgac cgcgcacctc ggcgacgcgg accaggcgcg ggcgagtcgc ggcggggtcc 2880
78 tcgccatctc acccgccgaa ggcatggagc tgttcgacgc agcgcgggac gggctcgtcg 2940
79 tcccgggtcaa cgtggacctg cgcaagacct cgcgcggcgg gacggtgccg cacctgctgc 3000
80 gcggcctggt ccgcccggga cggcagcagg cccgtccggc gtccactgtg gacaacggac 3060
81 tggccggggc actcgcgggg ctgcgcggcg cggagcagga ggcgctgctg ctcgacgtcg 3120
82 tccgcacgca ggtcgcgtg gtgtcggggc acgcccggcc ggaggcgtc cgcgcggaca 3180
83 cggcgttcaa ggacaccggc ttgactcgc tgacgtcgtt ggaactgcgc aaccggtgc 3240
84 gcgaggcgag cgggtgaag ctgcccgcga cgtcgtctt cgactaccgc acgcccgtcg 3300
85 cgtggcccc ctacctcgt gacgaattcg gcgacacggt ggcaacaact ccggtggcca 3360
86 cgcgcgcgc agcggacgcc ggcgagccga tcgccatcgt cggcatggcg tgcggctgc 3420
87 cgggcgggg caccgatccc gaaggcctgt ggcgcctggt gcgcgacggc ctogaagggc 3480
88 tgtctccctt ccccgaggac cggggctggg acctggagaa cctgttcgac gacgaccccg 3540
89 accgctccgg caccgctac accagccggg gcgggttctt cgacggcgcc ggctgttcg 3600
90 acgcccgtt ctccgggatt tcgcgcgcg aggcgctggc catggacccg cagcagcggc 3660
91 tgctgctcga ggcggcctgg gaagccctcg aaggcaccgg tgtcgacccg ggtcgttga 3720
92 agggcgccga cgtcggggtg ttgcgcgggg tgtccaacca gggctatggg atgggcgcgg 3780
93 atccggccga actggcgggg tacgcgagca cggcgggcgc ttcgagcgtc gtctcggggc 3840
94 gagtctcgt cgtcttcggg ttgaaggac cggcggtcac gatcgacacg gcttgctcgt 3900
95 cgtcgtggt ggcgatgcac ctggccgggc aggcgtgcg gcagggcgag tgctcgatgg 3960
96 cctggccgg tggcgtcac gtgatggga cgcgcggcac gttcgtggag ttgcggaagc 4020
97 agcgcggcct ggcgcgcgac ggccggtgca aggcctacgc cgaaggcgcg gacggcacgg 4080
98 gctgggcca gggcgtcggg gtgcgtcgtg tggagcggct gtcggtggcg cgcgagcgcg 4140
99 ggcaccgggt gctggccgtg ctgcgcggca gcgcggtcaa ctccgacggc gcgtccaacg 4200
100 gcctgaccgc cccaacggg cgtcgcagc aacgggtgat ccgcccggcc ctggccggcg 4260
101 ccggcctcga accgtccgat gtggacatcg tggagggca cggcaccggg acggcgtcg 4320
102 gcgacccgat cgaggcgcag gccctgctgg ccacctacgg caaggaccgc gaccggaga 4380
103 cgcgcttggt gctggggctg gtgaagtcga acttcggcca cgcgagtcg gcggccggcg 4440

```

RAW SEQUENCE LISTING

DATE: 07/11/2002

PATENT APPLICATION: US/10/042,665A

TIME: 09:11:59

Input Set : A:\ptoms.txt

Output Set: N:\CRF3\07112002\J042665A.raw

```

104 tggccgggggt gatcaagatg gtgcaggcgc tgcgccacgg cgtcatgccg cccaccctgc 4500
105 acgtggaccg gccaccagc caggtcgact ggtccgcggg ggccgtcgaa gtgctgaccg 4560
106 aggcacggga gtggccgcgg aacggccgtc cgcgcggggc cggggtgtcc tcgttcggga 4620
107 tcagcggcac gaacgcccac ctgatcatcg aagaagcacc ggccgagcca cagcttgccg 4680
108 gaccaccgcc ggacggcggg gtgggtgcgc tggtcgtctc ggctcgcagc cccggtgcc 4740
109 tggccgggtca ggcgcgctcg ctggccacgt tctcggcgca cggggccctt tccgacgtcg 4800
110 ccggtgcgct gacgagccgc gccctgttcg gcgagcgcg gcgtcgtcgtg gcggattcgg 4860
111 ccgaggaagc ccgcgcgggt ctggggcgac tggccgcgg cgaagacgcg ccgggacctg 4920
112 tccgcggccg ggtgcccgcg tccggcctgc cgggcaagct cgtgtgggtg tccccggggc 4980
113 aggggacgca gtgggtgggc atgggcccgc aactcctcga agagtctccg gtgttcgccg 5040
114 agcggatcgc cgagtgtgcg gccgcgctgg agccgtggat cggtcgtcg ctgttcgacg 5100
115 tctcctgtgg cgacggtgac ctcgatcggg tcgatgtgct gcagcccgcg tgctttgcgg 5160
116 tgatggtcgg cttggccgcg gtgtggtcct cggccggggg ggtccccgat gcggtgctcg 5220
117 gccactccca gggtagatc gccgcggcgt gcgtgtcggg tgcgttgctg ctggaggatg 5280
118 cggcgaaggt ggttgccctg cgcagccagg ccacgcgcgc gaagctctcc ggccgcggcg 5340
119 ggatggcttc ggtcgccttg ggcgaagccg atgtggtgtc gcggctggcg gacggggctg 5400
120 aggtggctgc cgtcaacggt ccggcgctcc tggtgatcgc ggggggatgcc caggccctcg 5460
121 acgaaacgct ggaagcgctg tccgggtgcg gaatccgggc tcggcggggtg gcggtggact 5520
122 acgcctcgca caccgggcac gtcgaagaca tcgaagacac cctcgcggaa gcgctggccg 5580
123 ggatcgacgc ccgggcgcgg ctggtgccgt tctctccac cctcaccggc gagtggatcc 5640
124 gggacgaggg cgctcgtggac ggcggctact ggtacc 5676

```

127 <210> SEQ ID NO: 2

128 <211> LENGTH: 1875

129 <212> TYPE: PRT

130 <213> ORGANISM: Amycolatopsis mediterranei

132 <400> SEQUENCE: 2

```

133 Tyr Pro Val Phe Ala Thr Ala Phe Asp Glu Ala Cys Glu Gln Leu Asp
134   1           5           10           15
136 Val Cys Leu Ala Gly Arg Ala Gly His Arg Val Arg Asp Val Val Leu
137           20           25           30
139 Gly Glu Val Pro Ala Glu Thr Gly Leu Leu Asn Gln Thr Val Phe Thr
140           35           40           45
142 Gln Ala Gly Leu Phe Ala Val Glu Ser Ala Leu Phe Arg Leu Ala Glu
143           50           55           60
145 Ser Trp Gly Val Arg Pro Asp Val Val Leu Gly His Ser Ile Gly Glu
146   65           70           75           80
148 Ile Thr Ala Ala Tyr Ala Ala Gly Val Phe Ser Leu Pro Asp Ala Ala
149           85           90           95
151 Arg Ile Val Ala Ala Arg Gly Arg Leu Met Gln Ala Leu Ala Pro Gly
152           100          105          110
154 Gly Ala Met Val Ala Val Ala Ala Ser Glu Ala Glu Val Ala Glu Leu
155           115          120          125
157 Leu Gly Asp Gly Val Glu Leu Ala Ala Val Asn Gly Pro Ser Ala Val
158           130          135          140
160 Val Leu Ser Gly Asp Ala Asp Ala Val Val Ala Ala Ala Ala Arg Met
161 145           150           155           160
163 Arg Glu Arg Gly His Lys Thr Lys Gln Leu Lys Val Ser His Ala Phe
164           165           170           175
166 His Ser Ala Arg Met Ala Pro Met Leu Ala Glu Phe Ala Ala Glu Leu

```

RAW SEQUENCE LISTING

DATE: 07/11/2002

PATENT APPLICATION: US/10/042,665A

TIME: 09:11:59

Input Set : A:\ptoms.txt

Output Set: N:\CRF3\07112002\J042665A.raw

```

167          180          185          190
169 Ala Gly Val Thr Trp Arg Glu Pro Glu Ile Pro Val Val Ser Asn Val
170          195          200          205
172 Thr Gly Arg Phe Ala Glu Pro Gly Glu Leu Thr Glu Pro Gly Tyr Trp
173          210          215          220
175 Ala Glu His Val Arg Arg Pro Val Arg Phe Ala Glu Gly Val Ala Ala
176 225          230          235          240
178 Ala Thr Glu Ser Gly Gly Ser Leu Phe Val Glu Leu Gly Pro Gly Ala
179          245          250          255
181 Ala Leu Thr Ala Leu Val Glu Glu Thr Ala Glu Val Thr Cys Val Ala
182          260          265          270
184 Ala Leu Arg Asp Asp Arg Pro Glu Val Thr Ala Leu Ile Thr Ala Val
185          275          280          285
187 Ala Glu Leu Phe Val Arg Gly Val Ala Val Asp Trp Pro Ala Leu Leu
188          290          295          300
190 Pro Pro Val Thr Gly Phe Val Asp Leu Pro Lys Tyr Ala Phe Asp Gln
191 305          310          315          320
193 Gln His Tyr Trp Leu Gln Pro Ala Ala Gln Ala Thr Asp Ala Ala Ser
194          325          330          335
196 Leu Gly Gln Val Ala Ala Asp His Pro Leu Leu Gly Ala Val Val Arg
197          340          345          350
199 Leu Pro Gln Ser Asp Gly Leu Val Phe Thr Ser Arg Leu Ser Leu Lys
200          355          360          365
202 Ser His Pro Trp Leu Ala Asp His Val Ile Gly Gly Val Val Leu Val
203          370          375          380
205 Ala Gly Thr Gly Leu Val Glu Leu Ala Val Arg Ala Gly Asp Glu Ala
206 385          390          395          400
208 Gly Cys Pro Val Leu Glu Glu Leu Val Ile Glu Ala Pro Leu Val Val
209          405          410          415
211 Pro Asp His Gly Gly Val Arg Ile Gln Val Val Val Gly Ala Pro Gly
212          420          425          430
214 Glu Thr Gly Ser Arg Ala Val Glu Val Tyr Ser Leu Arg Glu Asp Ala
215          435          440          445
217 Gly Ala Glu Val Trp Ala Arg His Ala Thr Gly Phe Leu Ala Ala Thr
218          450          455          460
220 Pro Ser Gln His Lys Pro Phe Asp Phe Thr Ala Trp Pro Pro Pro Gly
221 465          470          475          480
223 Val Glu Arg Val Asp Val Glu Asp Phe Tyr Asp Gly Phe Val Asp Arg
224          485          490          495
226 Gly Tyr Ala Tyr Gly Pro Ser Phe Arg Gly Leu Arg Ala Val Trp Arg
227          500          505          510
229 Arg Gly Asp Glu Val Phe Ala Glu Val Ala Leu Ala Glu Asp Asp Arg
230          515          520          525
232 Ala Asp Ala Ala Arg Phe Gly Ile His Pro Gly Leu Leu Asp Ala Ala
233          530          535          540
235 Leu His Ala Gly Met Ala Gly Ala Thr Thr Thr Glu Glu Pro Gly Arg
236 545          550          555          560
238 Pro Val Leu Pro Phe Ala Trp Asn Gly Leu Val Leu His Ala Ala Gly
239          565          570          575

```

RAW SEQUENCE LISTING

DATE: 07/11/2002

PATENT APPLICATION: US/10/042,665A

TIME: 09:11:59

Input Set : A:\ptoms.txt

Output Set: N:\CRF3\07112002\J042665A.raw

```

241 Ala Ser Ala Leu Arg Val Arg Leu Ala Pro Ser Gly Pro Asp Ala Leu
242          580          585          590
244 Ser Val Glu Ala Ala Asp Glu Ala Gly Gly Leu Val Val Thr Ala Asp
245          595          600          605
247 Ser Leu Val Ser Arg Pro Val Ser Ala Glu Gln Leu Gly Ala Ala Ala
248          610          615          620
250 Asn His Asp Ala Leu Phe Arg Val Glu Trp Thr Glu Ile Ser Ser Ala
251 625          630          635          640
253 Gly Asp Val Pro Ala Asp His Val Glu Val Leu Glu Ala Val Gly Glu
254          645          650          655
256 Asp Pro Leu Glu Leu Thr Gly Arg Val Leu Glu Ala Val Gln Thr Trp
257          660          665          670
259 Leu Ala Asp Ala Ala Asp Asp Ala Arg Leu Val Val Val Thr Arg Gly
260          675          680          685
262 Ala Val His Glu Val Thr Asp Pro Ala Gly Ala Ala Val Trp Gly Leu
263          690          695          700
265 Ile Arg Ala Ala Gln Ala Glu Asn Pro Asp Arg Ile Val Leu Leu Asp
266 705          710          715          720
268 Thr Asp Gly Glu Val Pro Leu Gly Arg Val Leu Ala Thr Gly Glu Pro
269          725          730          735
271 Gln Thr Ala Val Arg Gly Ala Thr Leu Phe Ala Pro Arg Leu Ala Arg
272          740          745          750
274 Ala Glu Ala Ala Glu Ala Pro Ala Val Thr Gly Gly Thr Val Leu Ile
275          755          760          765
277 Ser Gly Ala Gly Ser Leu Gly Ala Leu Thr Ala Arg His Leu Val Ala
278          770          775          780
280 Arg His Gly Val Arg Arg Leu Val Leu Val Ser Arg Arg Gly Pro Asp
281 785          790          795          800
283 Ala Asp Gly Met Ala Glu Leu Thr Ala Glu Leu Ile Ala Gln Gly Ala
284          805          810          815
286 Glu Val Ala Val Val Ala Cys Asp Leu Ala Asp Arg Asp Gln Val Arg
287          820          825          830
289 Val Leu Leu Ala Glu His Arg Pro Asn Ala Val Val His Thr Ala Gly
290          835          840          845
292 Lys Val Phe Ala Pro Lys Val Thr Ala Ala Asn His Leu Asp Glu Leu
293          850          855          860
295 Thr Arg Glu Leu Asp Leu Arg Ala Phe Val Val Phe Ser Ser Ala Ser
296 865          870          875          880
298 Gly Val Phe Gly Ser Ala Gly Gln Gly Asn Tyr Ala Ala Ala Asn Ala
299          885          890          895
301 Tyr Leu Asp Ala Val Val Ala Asn Arg Arg Ala Ala Gly Leu Pro Gly
302          900          905          910
304 Thr Ser Leu Ala Trp Gly Leu Trp Glu Gln Thr Asp Gly Met Thr Ala
305          915          920          925
307 His Leu Gly Asp Ala Asp Gln Ala Arg Ala Ser Arg Gly Gly Val Leu
308          930          935          940
310 Ala Ile Ser Pro Ala Glu Gly Met Glu Leu Phe Asp Ala Ala Pro Asp
311 945          950          955          960
313 Gly Leu Val Val Pro Val Lys Leu Asp Leu Arg Lys Thr Arg Ala Gly

```

VERIFICATION SUMMARY

DATE: 07/11/2002

PATENT APPLICATION: US/10/042,665A

TIME: 09:12:00

Input Set : A:\ptoms.txt

Output Set: N:\CRF3\07112002\J042665A.raw